

Genetic map of the human chromosome 10p15.5 region. The top part shows a chromosome with markers D13S283, D13S1243, ATP1A1, D13S787, D13S232, D13S1275, and a centromere. A 1 cM scale bar is provided. The bottom part shows a detailed map of the 11,487 bp ORF region, including markers M, S, a triangle, and polyA, with a 2 kb scale bar. A dashed line connects the 235\_L\_20 marker on the chromosome to the detailed map.

Figure 1



M	1	2	3	4	5	6	M	7	8	9	10	11	12	13	14
M	1	2	3	4	5	6	M	7	8	9	10	11	12	13	14

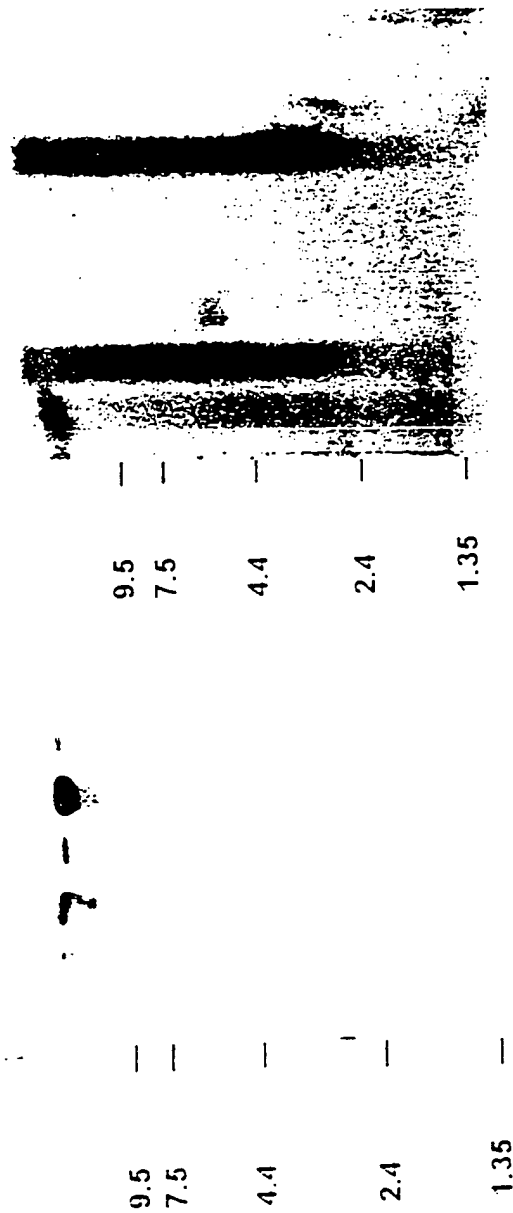
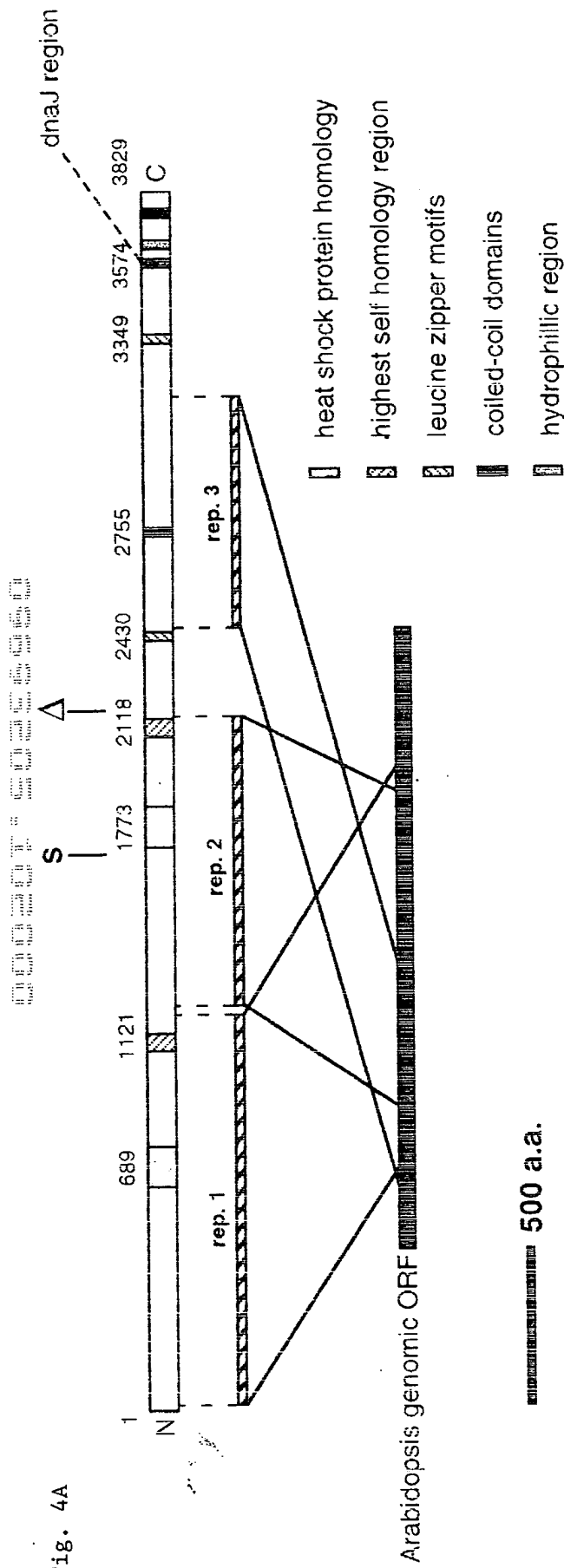
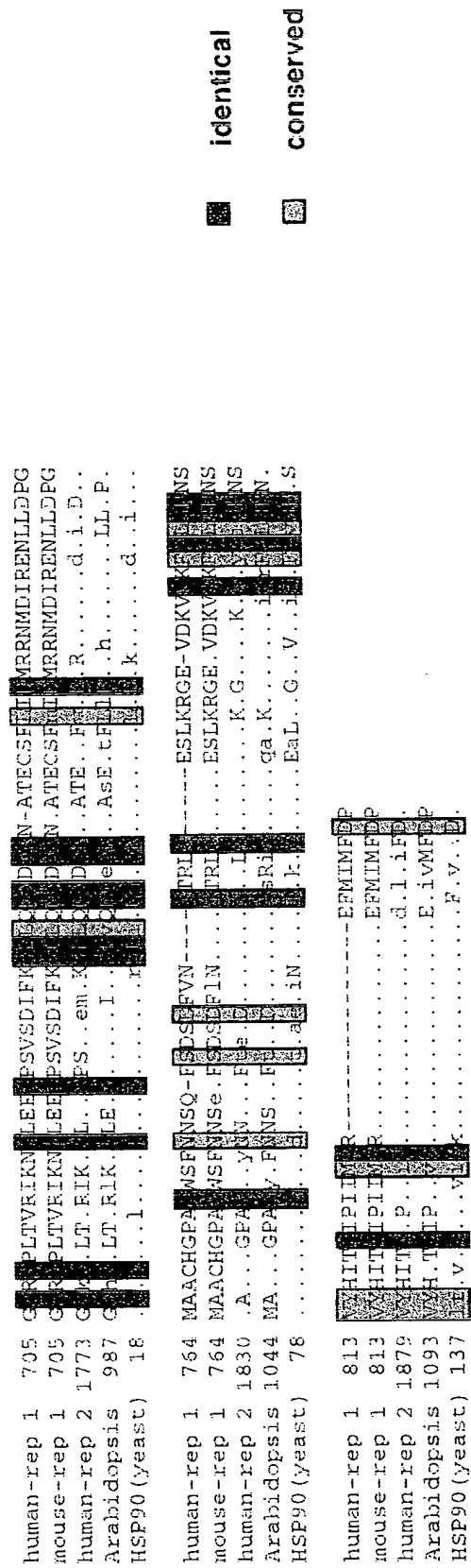


Fig. 4A



500 a.a.

Fig. 4B



(SEQ ID NO: 2)  
(SEQ ID NO: 4)

MNTFWPGSELIIVQWYFEDENRNHPSVSWL KQVWKNLYIHFSDDLTFDEMLIPRTILEE 60  
.....S.DKR...L.....L.N.  
GQTCVELIRLRIPSLVILDDSEEAQLPEFLADIVQKLGGLVKKLDASIQHPLIKKYIHS 120  
D.....V.....T.....I...R...T.....V.....  
PLPSAVLQIMEKMPLOKLCNQITSLLPHTKDALRKFLASLTDSSEKEKRIIQELAIFKRI 180  
.....I.....I.....A.....T.....T.....  
NHSSDQGISSYTKLKGCKVLHHTAKLPADLRSLISVIDSSDEATIRLANMLKIEQLKTTT 240  
.....D.....T.....V.....K.....  
CLKLVLDIENAFYSHEEVTQMLWVLENLSSLKNENPNVLEWLTPLKFIQISQEQMVSA 300  
...F....G....TQ.....I.....S...D...M....HM...GHV.A.  
GELFDPDIEVLKDLFCNEEGTYFPPSVFTSPDILHSLRQIGLKNEASLKEKDVQVAKKI 360  
.D.....R...Y...EAC...TI.....S.....R..  
EALQVGACPDQDVLLKKAKTLLLVLNKNHTLLQSSEGKMTLKKIKWVPACKERPPNYPGS 420  
.....SS.QN....M.....Q.....A.....  
LVWKGDLCLNLCAPPDMCDVGHAILIGSSPLVESIHVNLEKALGIFTKPSLSAVLKHFKI 480  
.....AA..V.V.....V.....Q..S.....TIN.....T  
VVDWYSSKTFSDDEDYQFQHILLEIYGFMDHDLNEGKDSFRALKFPWWVTGKKFCPLAQ 540  
.....T.....S.....K.....N.....  
VIKPIHDLDLQPYLHNVPKTMKFHQLFKVCGSIEELTSDHISMVIQKIYKSDQDLSEQ 600  
...T.....Y.....A.....V.....E...E  
ESKQNLHMLNIIRWLYSNQIPASPNTVPPIHHSKNPSKLIMKPIHECCYCDIKVDDLND 660  
.....M.....Y..R...V.....  
LLEDVPEPIILVHEDIPMKTAEWLKVPCSTRLINPENMGFEQSGQREPLTVRIKNILEE 720  
.....  
YPSVSDIFKELLQNAADDANATECSFLIDMRNMDIRENLLDPGMAACHGPALWSFNNSQF 780  
.....M.....E..  
SDSDFVNITRLGESLKRGEVDKVGKFGFGNSVYHITDIPIMSREFMIMFDPNINHISK 840  
.....L.....  
HIKDKSNPGIKINWSKQKRLRKFPNQFKPFIDVFGCQLPLTVEAPYSYNGTLFRLSFRT 900  
.....R.....A.....  
QQEAKVSEVSSTCYNTADIYSLVDEFSLCGHRLLIIFTQSVKSMYLYLKIEETNPSLAQD 960  
.....N.....  
TVIIKKSCSSKALNTPVLSVLKEAAKLMKTCSSSNKKLPSDEPKSSCILQITVEEFHHV 1020  
.I.....V.P....A.....T.V.....  
FRRIADLQSPLFRGPDDPAALFEMAKSGQSKKPSDELSCQKTVECTTWLLCTCMDTGEAL 1080  
.....T.....P...D....I.....  
KFSLSSESGRRLGLVPCGAVGVQLSEIQDQKWTVPKHIGEVFCYLPLRIKTGLPVHINGCF 1140  
....N.....L.H.T.E.....I.....  
AVTSNRKEIWKTDTKGRWNTTFMRHVIVKAYLQVLSVLRDLATSGELMDYTYAVWPDPD 1200  
.....A.....IG...T.....  
LVHDDFSVICQGFYEDIAHGKGKELTKVFSGDSTWVSMKNVRFLDDSIKRRDVGSAAFK 1260  
.....K.....R.....M.....Q.K.....  
IFLKYLKKTGSKNLCAVELPSSVKLGFEAGCKQILLENTFSEKQFFSEVFFPNIQIEIA 1320  
.....A.....  
ELRDPLMIFVLNEKVDEFSGVLRVTPCIPCSLEGHPLVLPRLIHPEGRVAKLFDIKDGR 1380  
.....N.....L.....I.....V.....T.....  
FPYGSTQDYLNPIILIKLVQLGKAKDDILWDDMLERAVSVAEINKSDHVAACLRSILLS 1440  
.....M.....E.....A.....

Figure 5A

LIDEKLIKIRDPRAKDFAAKYQTIRFLPFLTKPAGFSLDWKGNFSPKPMFAATDLYTAEH 1500  
 .....K.....P.....E.....I.....Y  
 QDIVC.....ILNENSHSFRGCGSVSLAVKEFLGLLKKPTVD.....QLKEVAKSVDDGITL 1560  
 .....Q.....  
 YQENITNACYKYLHEALMQNEITKMSIIDKLKPFSSFILVENAYVDSEKVSFHLNFEAAPY 1620  
 .....VL...MA.AT..E.....C.....V..E.....  
 LYQLPNKYKNNFRELFTVGVRSCTVEDFALVLESIDQERGKQITEENFQLCRRRISE 1680  
 .....S.....F.....K.....  
 GIWSLIREKKQEFCEKNYKILLPDTNMLLPAKSLCYNDPCWIKVKDITTVKYCHADIPR 1740  
 .....R.....L.....S.....  
 EVAVKLGAVPKRHKALERYASNVCFTTLGTEFGQKEKLTSLIKSILNAYPSEKEMLKELL 1800  
 .....I.....I.....A.....  
 ONADDAKATEICEVFDPROHPVDRIFDCKWAPLQGPALCVYNNQPFTEDEVRGIONLGKG 1860  
 .....  
 TRKGNPYKTGQYVGIGFNSVYHITDCPSFISGNDILCIFDPHARYAPGATSISPGRMFRDL 1920  
 .....C..H.....G.....V.....  
 DADFRTQFSDVLDLYLGTGHKLDNCTMFRFPLRNAEMAKVSEISSVPASDRMVQNLLDKL 1980  
 .....N.....Q.....S.....  
 RSDGAELLMFLNHMEKISICEIDKSTGALNVLYSVKGGKITDGDRLKQKFHASVIDSVTK 2040  
 .....A..G.....  
 KRQLKDIPVQQITYTMDTEDSEGNLTWILICNRSGFSSMEKVSXSVISAHKNQDITLFP 2100  
 .....  
 GGVAACITHNYKKPHRAFCFLPLSLETGLPFHVNGHFALDSARRNLWRDDNGVGRSDWN 2160  
 .....  
 NSLMTALIAPAYVELLIQLKKRYFPGSDPTLSVLQNTPIHVVKDTLKKFLSFFPVNRLDL 2220  
 .....  
 QPDLYCLVKALYNCIHEDMKRLLPVVRAPNIDGSDLHSAVITWINMSTSNKTRPFFDNL 2280  
 .....S.....  
 LQDELQHLKNADYNITTRKTVAENVYRLKHLLEIGFNLVYNCDATANLYHCLIDADIPV 2340  
 .....V.....  
 SYVTPADIRSFMTFSSPDNTCHIGKLPCLRLQNTNLKLFHSLKLLVDYCFKDAEENEIEV 2400  
 .....V.....S.F..  
 EGLPLLITLDSVLQTFDAKRPKFLTTYHELIPSRKDLFMTNTLYLKYSNILLNCKVAKVFD 2460  
 .....I..G.....SV.....  
 ISSFADLLSSVLPREYKTKSCTKWKNDFASESWLKNAWHFISESVSKEDQEETKPTFDI 2520  
 .....N.A.....TD...P..A..V  
 VVDTLKDWallPGTKFTVSANQLVVEGDVLLPLSLMHIAVFPNAQSDKVFHALMKAGCI 2580  
 I..I.....TS.....I.....  
 QLALNKICSKDSAFVPLLSCHTANIESPTSILKALHYMVQTSTFRAEKLVENDFEALLMY 2640  
 .....L.....D..A.....V.....T...M.....  
 FNCNLNHLMSQDDIKILKSLPCYKSISGRYVSIGKFGTCYVLTKSIPSAEVEKWTQSSSS 2700  
 .....S.....M..A.....  
 AFLEEKIHLKELYEVIGCVFVDDLEVLKHLLPKIENLSYDAKLEHLIYLNRL SSAEEL 2760  
 .....V.....L.....A.I..P  
 SEIKEOLFEEKLESLLIIH DANSRLKQAKHFYDRTVRVFEVMLPEKLFIPNDFFKLEQLI 2820  
 .....N.....KE.....V..  
 KPKNHVTFMTSWVEFLRNIGLKYILSQQQLLQFAKEISVRANTENWSKETLQNTVDILLH 2880  
 ....QAA.....A.....S.....

Figure 5B

HIFQERMDLLSGNFLKELSLIPFLCPERAPAEFIRFHPQYQEVNGTLPLIKFNGAQVNP 2940  
 .....Y.....  
 FKQCDVLQLLWTSCPIPEKATPLSIKEQEGSDLGPEQLEQVLNMLNVNLDPPDKVIN 3000  
 .....A.....  
 NCRNICNITTLDEEMVKTRAKVLRISIYEFLSAEKREFRQLRGVAFVMVEDGWKLLKPEE 3060  
 .....  
 VVINLEYESDFKPYLYKLPLELGTFFHQLFKHLGTEDIISTKQYVEVLSRIFKNSEKQLD 3120  
 .....A.....S.....  
 PNEMRTVKRVVSGLFRLQNDVSVKVRSDLENVRDLALYLPQDGRIVKSSILVFDDAPHY 3180  
 .....K.....A.....K.....  
 KSRIQGNIGVQMLVDLSQCYLGKDHGFHTKLIIMLFPQKLRPRLSSILEEQLEETPKVC 3240  
 .....  
 QFGALCSLQGRQLQLLSSEQFITGLIRIMKHENDNAFLANEKAIKALREGLKVSCF 3300  
 .....  
 EKLQTTLRVKGFNPIPHSRSETFAFLKRFGNAVILLYIQHSDSKDINFLLALAMTLKSAT 3360  
 .....  
 DNLI SDTSYLT IAMLGCNDIYRIGEKLD SLGVKYSSEPSKLELPMPGTPIPAEIHYTLLM 3420  
 .....S.....  
 DPMNVFYPGEYVGYLVDAEGGDIYGSYQPTYTYAIIVQEVEREDADNSSFLGKIYQIDIG 3480  
 .....T.....  
 YSEYKIVSSLDLYKFSRPEESSQSRDSAPSTPTSPTEFLTPLGRLSIPPLFSGRESHKT-S 3540  
 .....D...N...T.....K....SP..  
 SKHQSPKKLVNSLPEILKEVTSVVEQAWKI [REDACTED] 3600  
 T..H..R....A.....  
 [REDACTED] STSASRFQSDKYSFQRFYT SWNOEATSHK 3660  
 .....  
SERCOONKEK C PPSAGQYTSQRFFVPPTFKSVGNPVEARRWLRQARANFSAARNDLHKNA 3720  
 .....S.....  
 NEWVCFKCYLSTKLALIAADYAVRGKSDKDVKP TALAOKIEEYSOOLEGLTNDVHTLEAY 3780  
 .....  
 GVDSLKTRYPDLLPFPQIPNDRFTSEVAMRVMECTACIIKLENFMQQKV 3830  
 .....I.....

Figure 5C

**Table 1 ESTs identified by sample-sequencing of the ARSACS critical interval**

BAC clone	GenBank #	UniGene	Identity <sup>a</sup>	Tissue Source
235_1_20	AA987300	U15129092	221/230	neuroendocrine lung carcinoids
235_1_20	AA476635*		249/296	total fetus
235_1_20	AI351876*		272/335	melanocyte, fetal heart, pregnant uterus (pool)
235_1_20	W25994	U15163732	447/464	retina
235_1_20	AI377467	U15163732	257/263	total fetus
235_1_20	AA601007		488/491	schwannoma tumor
235_1_20	AA324964		214/234	cerebellum
235_1_20, 206_1_1	AA897178 <sup>b</sup>		238/238	fetal lung, testis, and B-cell
235_1_20, 206_1_1	R17106 <sup>b</sup>	U15188560	747/784	brain, adipose tissue
235_1_20, 206_1_1	AB018273 <sup>b</sup>	U15159492	4318/4318	multiple tissue types including brain, CNS, and whole embryo

a: Number of homologous nucleotides between the BAC subclone sequence and the GenBank sequence

b: AA897178, R17106, and AB018273 are all contained within *spastin*.

\* AA476635 and AI351876 were homologous to non-overlapping portions of the same M13 subclone sequence.



# Table 2 Primers for PCR amplification of the human *spastin* ORF

Primer set	Forward primer	Reverse primer	Product size (bp)
1 (SEQ ID NO: 21)	CCTTCCAGTACTGTGTTATTTGTGAG	CAAGAACTTCCTCAGGGGCATC	(SEQ ID NO: 22) 603
2 (SEQ ID NO: 23)	GATGCA'TCTATATACAAATCCGCT'	GGTGGGAATAGGTTCCTTC	(SEQ ID NO: 24) 581
3 (SEQ ID NO: 25)	AAAAATGAGAA'TCCAAATGTGCT'	GCACTAAGGCTAGGTTTGTGAAG	(SEQ ID NO: 26) 592
4 (SEQ ID NO: 27)	GCTCCTCAC'TTCCTCTTGTG	CGTGAA'TTGGCTTCATGATAA	(SEQ ID NO: 28) 602
5 (SEQ ID NO: 29)	AGCAATCAGATTCCAGCAAAGC	GATGGAA'TGTCAGTGATATGG	(SEQ ID NO: 30) 611
6 (SEQ ID NO: 31)	GGGAGAAAGTTGACAAAGTTGGA	CTTTGGTTCATCACATGGGAAG	(SEQ ID NO: 32) 624
7 (SEQ ID NO: 33)	TCCAAAGCATTTGAACACACCT'	CAGTCCCGTAAAGACATCAG	(SEQ ID NO: 34) 631
8 (SEQ ID NO: 35)	CAATGGGTGCTTTGCTGTAC	CGAAGAACTCCCGAGAACTCA	(SEQ ID NO: 36) 620
9 (SEQ ID NO: 37)	GCTGGCTGCAAAACAGATACTAC	GCAAAACATGGTTTCAGGGCTTA	(SEQ ID NO: 38) 604
10 (SEQ ID NO: 39)	CAAAACAA'TCCGCTTCCTTCCAT	ATTATTCGTCCGGCAAAGCTGA	(SEQ ID NO: 40) 651
11 (SEQ ID NO: 41)	TTCCGCGAACTTTTGAACCC	ACACAAGTGTCTGGCCCTTGC	(SEQ ID NO: 42) 625
12 (SEQ ID NO: 43)	GATGCAAAAGCGACAGAAATC	ATACAGCACATTTAGAGCTCCAGT	(SEQ ID NO: 44) 626
13 (SEQ ID NO: 45)	GCATCAGACAGAAATGGTCCAG	GCAATTCAACATA'TGCAGGAG	(SEQ ID NO: 46) 624
14 (SEQ ID NO: 47)	GTGAATGGCCACTTTGCACT	TGATATCAGCAGGGGTCACAT	(SEQ ID NO: 48) 648
15 (SEQ ID NO: 49)	ACCACACGCAAAACAGTAGCA	GCCATGCA'TTCTT'AAGCCCAAG	(SEQ ID NO: 50) 609
16 (SEQ ID NO: 51)	TGACATTTCCAGCTTTGCTGA	AGCGGCCACTGATGGA'TTTAT	(SEQ ID NO: 52) 631
17 (SEQ ID NO: 53)	AAATGATTTTGAGGCACTTTTG	TTCCACCCAGGATGTCATAAA	(SEQ ID NO: 54) 609
18 (SEQ ID NO: 55)	ACAGTAGACTAAAGCAAGCAAAGC	ATCAAGAGGAGGATCCAGGTI'	(SEQ ID NO: 56) 645
19 (SEQ ID NO: 57)	CATCCTGCCCTATTC'TCCAG	TAAAGCGCAAGGTCTCTGTACA	(SEQ ID NO: 58) 618
20 (SEQ ID NO: 59)	TGAGGGCAAAACAA'TAGATCC	TC'TGCTGTGGGGAATAGGATTT	(SEQ ID NO: 60) 612
21 (SEQ ID NO: 61)	GCAAAGCCCTAAGAGAAAGGATT	TGCTTTGAGAGCTTTCCTCAG	(SEQ ID NO: 62) 647
22 (SEQ ID NO: 63)	TGAAAGAGAAAGATGCTGACAATTC	GTAAGTCTGTCCGGCTGAAGG	(SEQ ID NO: 64) 654
23 (SEQ ID NO: 66)	CATCCCGATTTCAGTCAGACA	TTCTGTCTACAACACATTCAGA	(SEQ ID NO: 66) 638

Figure 7

LOCUS AF193557 11493 bp DNA ROD  
 DEFINITION Mus musculus saccin gene, complete cds.  
 ACCESSION AF193557  
 VERSION AF193557.1 GI:6907043  
 KEYWORDS  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 11493)  
 AUTHORS Engert, J.C., Berube, P., Mercier, J., Dore, C., Lepage, P., Ge, B.,  
 Bouchard, J.P., Mathieu, J., Melancon, S.B., Schalling, M.,  
 Lander, E.S., Morgan, K., Hudson, T.J. and Richter, A.  
 TITLE ARSACS, a spastic ataxia common in northeastern Quebec, is caused  
 by mutations in a new gene encoding an 11.5-kb ORF  
 JOURNAL Nat. Genet. 24 (2), 120-125 (2000)  
 MEDLINE 20120709  
 REFERENCE 2 (bases 1 to 11493)  
 AUTHORS Engert, J.C., Berube, P., Dore, C., Lepage, P., Ge, B., Hudson, T.J. and  
 Richter, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-OCT-1999) Genome Centre, Montreal General Hospital,  
 1650 Cedar Ave., Montreal, QC H3G 1A4, Canada  
 FEATURES Location/Qualifiers  
 source 1..11493  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
mRNA <1..>11493  
 /product="saccin"  
CDS 1..11493  
 /note="molecular chaperone"  
 /codon\_start=1  
 /product="saccin"  
 /protein\_id="AAF31263.1"  
 /db\_xref="GI:6907044"  
 /translation="MNTFWPGRELVVQWYPFSEDKRHPSLSWLKMVWKNLYIHFSEDL  
 TLFDEMPLIPRTLLNEDQTCVELIRLRIPSVVILDDTEAQLPEFLADIVQKLGGIVL  
 KRLDTSIQHPLVKKYIHSPLPSAILQIMEKIPLQKLCNQIASLLPTHKDALKFLASL  
 TDTSEKEKRRIQELTIFKRINHSSDQGISSYTKLKGCKVLDHTAKLPTDLRLSVSVID  
 SSDEATIRLANMLKIEKLKTTSCCLKFVLKDIGNAFYTQEEVTQLMLWILENLSSLKNE  
 NSNVLDWLMPLKFIHMSQGHVVAAGDLFDPDIEVLRDLFYNEEEACFPPTIFTSPDIL  
 HSLRQIGLKNESLKEKDVVQVARKIEALQVSSCQNQDVLMMKAKTLLLVLNKNQTLL  
 QSSEGKMAKKIKWVPACKERPPNYPGSLVWKGDLNLCAPPDMCDAAHAVLVGSSLP  
 LVESVHVNLQALSIFTKPTINAVLKHFKTVVDWYTSKTFSDDEDYYQFHILLEIYGF  
 MHDHLSSEGKDSFKALKFPVWWTGKNFCPLAQAVIKPTHDLDLQPYLYNVPKTMKFHQ  
 LFKACGSIEELTSDHISMVIQKVYLKSDQELSEEESKQNLHMLNIMRWLYSNQIPAS

Figure 8A

PNTVPVIYHSRNP SKLVMKPIHECCYCDIKVDDLNDLLED SVEPIILVHEDIPMKTAE  
WLKVPCLSTRLINPENMGFEQSGQREPLTVRIKNILEEYPSVSDIFKELLQNADDANA  
TECSFMIDMRNMDIRENLLDPGMAACHGPALWSFNNSEFSDSDFLNITRLGESLKR  
EVDKVGKFG LGFNSVYHITDIPIIMSREFMIMFDPNINHISKHIKDRSNPGIKINWSK  
QQKRLRKFPNQFKPFIDVFGCQLPLAVEAPYSYNGTLFRLSFRTQQEAKVSEVSSTCY  
NTADIYSLVDEFSLCGHRLLIIFTQSVNSMYLKYLKIEETNP SLAQDTIIKKKVCPSK  
ALNAPVLSVLKEAAKLMKTCSSSNKKLPTDVPKSSCILQITVEEFHHVFRRRIADLQSP  
LFRGPDDDPATLFEMAKSGQSKKPSDEL PQKTVDCTTWLICTCMDTGEALKFSLNESG  
RRLGLVPCGAVGVLLHETQE QKWTVKPHIGEVFCYLPLRIKTGLPIHINGCFAVTSNR  
KEIWKTDTKGRWNTTFMRHVIVKAYLQALSVLRLDAIGGELTDYTYAVWPDPLVHD  
DFSVICKGFYEDIAHGKGKELTRVFS DGSMWVSMKNVRFLDDSILQRKDVGSAAFKIF  
LKYLKKTGSKNLCAVELPSSVKAGFEEAGCKQILLENTFSEKQFFSEVFFPNIQEIEA  
ELRDPLMNFVLNEKLDEFSGILRVTPCVPCSLEGHPLVLP SRLIHPEGRVAKLFDTKD  
GRFPYGSTQDYLNPIILIKLVQLGMAKDDILWDDMLERAESVAEINKSDHAAACLRSS  
ILLSLIDEKLKIKDPRAKDFAAKYQTI PFLPFLTKPAGFSLEWKGNSFKPETMFAATD  
IYTAEYQDIVCLLQPILNENSHSFRGCGSVSLAVKEFLGLLKKPTVDLVINQLKQVAK  
SVDDGITLYQENITNACYKYLHEAVLQNE MAKATIIIEKLKPF CFILVENVYVESEKVS  
FHLNFEEAAPYLYQLPNKYKNNFREL FESVGRQSFTVEDFALVLESIDQERGKKQITE  
ENFQLCRRIISEGIWSLIREKRQEFCEKNY GKILLPDTNLLLLPAKSLCYND CPWIKV  
KDSTVKYCHADIPREVAVKLGAI PKRHKALERYASNICFTALGTEFGQKEKLT SRIKS  
ILNAYPSEKEMLKELLQNADDAKATEICFVFDPRQHPVDRI FDDKWAPLQGPALCVYN  
NQPFTEDDVRGIQNLGKGTKEGNPCKTGHY GIGFNSVYHITDCPSFISGNDILGIFDP  
HARYAPGATSVSPGRMFRDL DADFRTQFSVDLDLYLGNHFKLDNCTMFRFPLRNAEMA  
QVSEISSVPSSDRMVQNLLDKLRSDGAELLMFLNHMEKISICEIDKATGGLNVLYSVK  
GKITDGDRLKRKQFHASVIDSVTKKRQLKDIPVQQITYTMDTEDSEGNLTTWLICNRS  
GFSSMEKVSKSVISAHKNQDITLFP RGGVAACITHNYKKPHRAFCFLPLSLETGLPFH  
VNGHFALDSARRNLWRDDNGVGRSDWNNSLMTALIAPAYVELLIQLKKRYFPGSDPT  
LSVLQNTPIHVVKDTLKKFLSFFPVNRLDLQPDLYCLVKALYSCIHEDMKRLLPVVRA

Figure 8B

PNIDGSDLHSAVITWINMSTSNKTRPFFDNLQDELQHLKNADYNITTRKTVAENVY  
RLKHLLEIGFNLVYNCDETANLYHCLVDADIPVSYVTPADVRSFLMTFSSPDTNCHI  
GKLPCRLLQQTNLKLFHSLKLLVDYCFKDAEESFEVEGLPLLITLDSVLQIFDGKRPK  
FLTTHYELIPSRKDLFMNTLYLKYSSVLLNCKVAKVFDISSFADLLSSVLPREYKTKN  
CAKWKDNFASESWLKNAWHFISESVSVTDDQEEPKPAFDVIVDILKDWALLPGTKFTV  
STSQLVVPEGDVLIPLSLMHIAVFPNAQSDKVFHALMKAGCIQLALNKICSKDSALVP  
LLSHTANIDSPASILKAVHYMVQTSTFRTEKLMENDFEALLMYFNCNLSHLMSQDDI  
KILKSLPCYKISGRYMSIAKFGTCYVLTKSIPSAEVEKWTQSSSSAFLEEKVHLKEL  
YEVLCVVPDDLEVYLKHLPLKIENLSYDAKLEHLIYLNRLASIEEPSEIKEQLFEK  
LESLLIHDANNRLQAKHFYDRTVRVFEVMLPEKLFIPKEFFKKLEQVIKPNQAAF  
MTSWVEFLRNIGLKYALSQQQLQFAKEISVRANTENWSKETLQSTVDILLHHIFQER  
MDLLSGNFLKELSLIPFLCERAPAHEYIRFHPQYQEVNGTLPLIKFNGAQVNPFKKQC  
DVLQLLWTSCPILPEKATPLSIKEQEGSDLAPQEQLQVLNMLNVNLDPPDKVINNC  
RNICNITTLDEEMVKTRAKVLRISIYEFLSAEKREFRFQLRGVAFVMVEDGWKLLKPEE  
VVINLEYEADFKPYLYKLPLELGTFFHQLFKHLGTEDIISTKQYVEVLSRIFKSSEGKQ  
LDPNEMRTVKRVVSGFLFKSLQNDSVKVRSDLENARDLALYLPQDGKLVKSSILVFDD  
APHYKSRIQGNIGVQMLVDLSQCYLGKDHGFHTKLIMLFPQKLRPRLLSSILEEQLDE  
ETPKVCQFGALCSLQGRLLQLLSSEQFITGLIRIMKHENDNAFLANEKKAIRLCKALR  
EGLKVSCFEKLQTTLRVKGFNPPIPHSRSETFAFLKRFNGAVILLYIQHSDSKDINFL  
ALAMTLKSATDNLISDTSYLIAMLCNDIYRISEKLDLGVKYDSSEPSKLELPMPT  
PIPAEIHYTLLMDPMNVFYPGEYVGYLVDAEGGDIYGSYQPTYTYAIIVQEVEREDAD  
NTSFLGKIYQIDIGYSEYKIVSSLDLYKFSRPDESSQNRDSAPTTPPTSPTEFLTPGLR  
SIPPLFSGKESHKSPSTKHHSRKLKVNALPEILKEVTSVVEQAWKLPESERKKIIRR  
LYLKWHPDKNPENHDIANEVFKHLQNEINRLEKQAFDQADRASRRTFSTSASRFQS  
DKYSFQRFYTSWNQEATSHKSERQQQSKEKCPPSAGQTYQRFFVPPTFKSVGNPVEA  
RRWLRQARANFSAARNDLHKANANWVCFKCYLSTKLALIAADYAVRGKSDKDVKPTAL  
AQKIEEYSQQLEGLTNDVHTLEAYGVDSLKTRYPDLLPFPQIPNDRFTSEVAMRVMEC  
TACIIIKLENFIQQKV"

Figure 8C

BASE COUNT 3599 a 2281 c 2387 g 3226 t  
ORIGIN

```

1 atgaatacat tctggcctgg tcgagagttg gtgggttcagt ggtatccatt tagtgaagac
61 aaacgtcacc catccctttc atggccttaag atgggtttgga agaattctcta tatacatttc
121 tcggaagatt tgactttatt tgatgagatg ccacttatcc ctagaactct actgaatgag
181 gaccagacgt gtgtggaact catcagactc aggatcccat cagtagtcat tttagatgat
241 gaaactgaag ctcagcttcc agaattctta gcagatattg tacaaaaact tggagggatt
301 gtcctgaaaa gactagatac ctctattcag catccacttg ttaaaaaata cattcattcc
361 ccactcccga gtgctatttt gcagataatg gagaagatac ctctacagaa gttgtgtaat
421 caaatagcat cattacttcc aaccacaaa gatgctctaa ggaagttttt ggccagctta
481 actgatacca gtgaaaaaga gaaaagaata attcaagaat tgacaatatt caaaagaatt
541 aatcactcat cagatcaagg gatttcctct tacacaaaat taaaaggatg taaagttttg
601 gatcataccg ccaagcttcc aacagatcta cggctatcag tttcagtaat agatagtagt
661 gatgaagcca ccattcgttt ggcaaacatg ttgaaaattg aaaaattgaa gactacaagc
721 tgtttaaagt ttgtttttaa agatattgga aatgcatttt atacacagga agaggtaaca
781 caacttatgc tttggatcct tgagaatcta tcctctctta aaaatgagaa ttcaaatgtg
841 cttgattggg taatgccact aaaattcatt catatgtccc agggacatgt ggtagcagct
901 ggtgatctct ttgatcctga tatagaagta ctaagggatc tcttttataa tgaagaagaa
961 gcttgtttcc cacctacaat ttttacctca ccagatatcc ttcactcttt gagacagatt
1021 ggcttaaaaa atgaatccag tctaaaagaa aaagatgttg tacaagtggc aagaaaaatt
1081 gaagctttac aggtcagttc ctgtcagaat caggatgttc tcatgaagaa agccaaaaca
1141 ctcttactgg tcttgaataa aaaccagaca ctcttgcaat cttctgaagg gaagatggca
1201 ttgaagaaaa tcaaatgggt tccagcctgc aaggaaagac ctccaaatta tcccggttcc
1261 ttagtctgga aaggggatct ctgtaatctt tgtgcacctc cagatatgtg tgatgcggca
1321 catgcagttc tagtaggctc ctcaacttct cttgttgaaa gtgtccatgt gaacctggag
1381 caggcgctca gcatcttcac aaagcctact atcaatgctg tcttaaaaca ctttaaaact
1441 gttgttgact ggtatacttc aaaaaccttt agtgatgaag attactatca gttccaacat
1501 attttgcttg aaatttatgg gttcatgcat gatcatctga gtgaagggaa ggattctttt
1561 aaagccttga agtttccatg gggttgact ggcaaaaact tttgtcctct tgcccaggct
1621 gtgataaagc caaccatga tctggatctt cagccttatt tatataatgt gcctaaaacc
1681 atggcaaaa tccaccatga gttcaaggct tgtggctcaa tagaagagt gacatcagat
1741 catatttcca tggtcattca gaaagtttat ctcaaaaagt accaggagt gagtgaagaa
1801 gaaagtaaac aaaatcttca tctcatgttg aatattatga gatggctcta tagcaatcag
1861 attccagcaa gccctaatac accagttcct atttatcaca gcagaaatcc ttccaaactt
1921 gtcatagaag caattcatga atgctgttat tgtgacatca aagttgatga cctcaatgac
1981 ttgcttgaag attcagtgga accaattatc ttggtacatg aagatatacc catgaaaact
2041 gcagaatggc taaaagttcc gtgccttagt acaagactga tcaatcctga aaacatgggg
2101 tttgagcagt cagggcaaag agagcctctt actgtaagga ttaaaaaat tttggaagaa
2161 tacccttccg tgtcagatat ttttaaagag ctacttcaaa atgctgatga tgcaaatgcc
2221 acagaatgca gcttcatgat tgatatgaga aggaatatgg acatacggga aaatctcctg
2281 gaccaggga tggcagcttg tcatggacct gctctgtggg cattcaacaa ttctgaattc
2341 tcagattcag atttctttaa cataacgagg ttaggagagt ctttaaaaag gggagaagtt
2401 gacaagggtg ggaaatttgg tcttggtttt aattctgtgt accacatcac tgacattccc
2461 atcattatga gcagagaatt tatgataatg tttgatccaa acataaatca tatcagcaaa
2521 cacattaaag atagatcgaa tcctggaatc aaaattaatt ggagtaagca gcagaaaaga
2581 cttaggaagt tccccacca gttcaaacca tttatagatg tatttggtct tcagttacct
2641 ttggctgttg aagctcctta cagctacaat ggaactcttt tccgactgtc ctttagaaca
2701 cagcaggaag caaaagttag tgaagttagc agtacttgct acaatactgc ggaattttac
2761 tccctagtgg atgaatttag tcttgttggg cacagactta tcatttttac tcagagtgtg
2821 aactcagtgt atttgaaata cttgaaaatt gaagaaacca atcctagctt agcacaagat
2881 acaatcataa ttaagaaaaa agtttgcccc tccaaagcat tgaatgcacc agttttaagt
2941 gtttttaaaag aagctgctaa actcatgaag acttgtagca gcagcaacaa gaagcttccc
3001 acggatgtgc caaagtcac ttgcattctt cagatcacag tcgaagaatt ccaccatgtg
3061 tttaggagga ttgctgactt acagtcacca ctatttcgag gtccagatga tgaccagct
3121 actctctttg aaatggctaa atctggccaa tcaaaaaagc catcagatga gttgccacaa
3181 aagacagtag attgtaccac atggccttata tgcacatgca tggatacagg agaagctctc
3241 aagttttcct tgaatgaaag tggagaagaa ttagggctgg ttccttgtgg ggcagtaggg

```

Figure 8D

3301 gttctcttgc atgaaaccca ggaacagaag tggaccgtga aaccacacat aggagaagtg  
3361 ttttgctatt tacctctacg aatcaaaaca ggggttgccaa ttcacatcaa tgggtgcttt  
3421 gctgttactt caaataggaa agaaatctgg aagacagata caaaaggctg atggaatacc  
3481 acattcatga ggcatgtcat tgtgaaagct tacttacaag ccctcagtggt cttacgggac  
3541 ctagccattg gtggtgagct gactgattat acttactatg cagtgtggcc tgatcctgat  
3601 ctagtctcatg atgacttctc tgtgatctgt aaaggathtt atgaagacat tgctcatggg  
3661 aaggggaagg agttgaccag agtcttctct gatgggtcta tgtgggtttc catgaagaat  
3721 gtgaggtttc tggatgactc tatacttcaa aggaaagatg ttggttcagc agccttcaag  
3781 atatttctga agtacctcaa gaaaacagga tccaaaaacc tctgtgctgt tgagcttctc  
3841 tcttcagtaa aagcaggatt tgaagaggct ggctgtaaag agatactgct ggaaaataca  
3901 ttttcagaga aacagttctt ttccagaagtc ttctttccta atatccagga aattgaagca  
3961 gaacttagag atcctctgat gaattttgtc ctaaatgaaa aacttgatga gttctcagga  
4021 attcttctgt ttacccttgg tgttcttctg tccttggagg gccatccttt ggttttgcct  
4081 tcaagattga tccatcctga aggacgagtt gcaaagtatt ttgatactaa agatggaaga  
4141 ttcccttatg gttccacaca ggattacctc aatcctatta tcttgattaa gctcgttcag  
4201 ttaggcatgg caaaagatga tattttgtgg gatgacatgc tagagcgtgc agagtctgta  
4261 gctgagatta ataaaagtga ccatgctgct gcctgcttaa ggagtagtat tctgctaagc  
4321 cttattgatg agaagctaaa aataaaggat cctagagcaa aggattttgc tgcaaaatat  
4381 caaacaattc ctttctctcc atttctaaca aagccagcag gtttttcttt agaatggaaa  
4441 gggaaacagct ttaagcctga aacctgthtt gcagcaactg acatttacac agctgaatat  
4501 caagatatag tctgtctttt gcaaccaatt cttaatgaaa attcccattc ctttagaggc  
4561 tgtggttcag tgtctttggc tgttaaggag tttttgggtt tactaaagaa gccaacagtt  
4621 gatctggtaa taaaccagtt gaagcaagtt gcaaaatcag ttgatgatgg cattacattg  
4681 taccaggaaa atatcaccaa cgcttgctac aaatacctcc atgaagcagt attgcagaat  
4741 gaaatggcca aggcaacaat tattgagaag ctaaagccat tttgtttcat tctagttgag  
4801 aatgtatatg ttgagtcaga aaaggthttc tttcacttga actttgaagc agcaccatac  
4861 ctttatcagt tacctaaca gtataaaaaa aatttccgtg agctttttga aagtgtgggt  
4921 gtgcgacagt catttactgt tgaagacttt gccttagthtt tggagctcat tgatcaagag  
4981 agaggaaaaa aacaaaatac agaagagaat ttccagcttt gccgacgaat aatcagtgaa  
5041 ggcactctga gtctcattag agaaaagaga caagaatttt gtgagaaaaa ttatggcaaa  
5101 atattactgc cagacactaa cctgctgctg ctccctgcta agtcattatg ctacaatgac  
5161 tgtccctgga taaaagtaaa ggactccact gtcaagtatt gccatgccga cataccccg  
5221 gaagtagctg taaaacttgg tgcaatacca aagagacata aagcattaga aagatatgca  
5281 tccaacatct gtttcacagc tctaggtaca gaatttgggc agaaagaaaa actgaccagc  
5341 agaattaaga gcattctcaa tgcctatcct tcagaaaagg aaatgctgaa agagcttctt  
5401 caaaatgctg atgatgcaaa ggccacagag atctgctttg tgtttgatcc tagacagcat  
5461 cctgttgacc gaatatttga tgataagtgg gccccactgc aagggccagc actgtgtgtt  
5521 tacaacaacc agccatttac agaagatgat gttagaggaa ttcagaatct tgggaaaggc  
5581 accaaagaag ggaatccttg caaaacagga cattatggaa tcggattcaa ttccgtttat  
5641 catattacag actgcccttc ttttatttct ggcaatgaca tcctgggtat ttttgatccc  
5701 catgccagat atgcaccagg agccacatca gttagccctg gacgcatgtt tagagatttg  
5761 gatgcagact ttagaaccca gttctcagat gttctagatc tgtacttggg aaaccacttt  
5821 aaactggaca attgtacaat gtttagattt cctctgcgta atgcagagat ggcacaagtt  
5881 tcagaaattt cttccgttcc atcatcagac agaattggtc agaactttt ggacaagtta  
5941 cgtctgatg gggcagaact tctaattgth ctaaccaca tggagaaaa atctatttgt  
6001 gaaatagata aggccacagg aggtctgaat gtgctctatt cagtaaaagg caagatcact  
6061 gatggagacc gattgaaaag gaagcaattc cagcctctg taattgacag tgttactaaa  
6121 aagagacagc tcaaggacat accagttcaa caaataacct acactatgga tactgaggat  
6181 tctgaaggaa atctgaccac atggctcatc tgtaatagat caggattttc aagtatggaa  
6241 aaagtatcca agagtgtaat atcagctcac aagaaccaag atatcaccct tttccacgt  
6301 ggtggagtag cagcctgcat tactcacaat tataaaaagc cccacagagc cttctgcttt  
6361 ctgctctct ctttggagac agggctgcca tttcatgtga atggccactt tgctctagat  
6421 tcagccagaa gaaacttgtg gcgtgatgat aatggggttg gtgttcgaag tgactggaat  
6481 aatagtttaa tgacagcatt aatagcacct gcatatgttg agttactaat ccagttaaaa  
6541 aaacggtatt tccctggttc tgaccaaca ttatcagtht cagttaacag gctggatctg  
6601 gtcgtaaaag acacattaaa gaagthttct tccttctttc cagttacag gattcatga  
6661 cagccggact tatattgctt agtaaaagca ctttacagtt gcattcatga agacatgaag

Figure 8E

6721 cgtcttttgc ctgttggttcg ggctccaaat attgatggct cagatttgca ctctgcagtc  
6781 ataattactt ggatcaatat gtctacttca aataaaacta gaccattttt tgataactta  
6841 ctacaggatg aattacagca ccttaaaaat gcagattata acatcacaaac tcgaaaaaca  
6901 gtcgcagaga atgtctacag actgaagcac ctgctcttag aaattgggtt caacttggtt  
6961 tataactgtg atgaaactgc taacctttac cattgccttg tagatgcaga tatccctgtc  
7021 agctatgtga ctctgtctga tgttaggtcc ttcttaatga ctttctcttc tcctgacact  
7081 aattgccata ttgggaagct gccttgctgt cttcagcaga ctaacctaaa actttttcac  
7141 agtttaaaac ttttagttga ttactgtttt aaagatgctg aagaaagtga gtttgaagtt  
7201 gagggactgc ccctactcat tacactggac agtgtcttgc agatttttga tggtaaacga  
7261 cccaagtttc taacaacata ccatgaatta attccatcgc gtaaagactt gtttatgaac  
7321 accttatact tgaaatacag tagtgttttg ttgaactgca aagttgcaaa agtgtttgac  
7381 atttccagct ttgctgactt actctcttct gtgttgcttc gtgagtacaa gacaaaaaac  
7441 tgtgcaaagt ggaaagacaa ttttgccagt gaatcttggc ttaagaacgc atggcatttt  
7501 atcagtgaat cagtaagtgt aacggatgat caggaagaac caaagccagc atttgatgtc  
7561 attgttgaca tccttaaaga ctgggcattg cttccaggaa caaagttcac tgtgtcaacc  
7621 agtcagcttg tggttcctga gggagacgtg ttgattcccc tgagcctcat gcacattgct  
7681 gtgttcccaa atgctcagag tgataaggtt tttcacgctc tgatgaaagc tggctgtatt  
7741 cagctggctt tgaacaaaat ctgctctaaa gacagcgcac tagttcctct gttgtcatgc  
7801 cacacagcaa acatagatag ccctgcaagc atcttgaagg ctgtgcatta tatggttcag  
7861 acgtcaacat ttagaactga aaaactaatg gaaaatgact ttgaagcact tttgatgtat  
7921 ttcaactgta atttgagtca cttgatgtcc caagatgaca taaaaatttt aaagtcacct  
7981 ccattgctaca aatccatcag tggccgctat atgagcattg caaaatttgg aacgtgctat  
8041 gtgcttacca aaagtattcc ttcagctgaa gtggaaaaat ggacacagtc atcctcttcc  
8101 gcgtttcttg aagaaaaggt gcatttaaaa gaactctatg aggtgcttgg ctgtgtgcca  
8161 gtagatgatc tggaggtgta tttgaaacat cttctgccaa aaattgaaaa tctctcttat  
8221 gatgcaaagt tggagcacct gatttatctg aagaatagac tggcaagcat cgaggaaccg  
8281 tcagagatta aggagcaact ttttgaaaaa ctggaaagct tattgattat ccacgatgcc  
8341 aacaactcag taaagcaagc aaaacatttc tatgacagaa ctgtgagagt ttttgaagtt  
8401 atgcttctcg aaaaattggt tattcctaag gatttcttta aaaaattgga acaagtaatc  
8461 aaacctaaaa atcaagctgc atttatgacg tcctgggtgg aattcttgag aaatattgga  
8521 ctgaagtacg cgctctccca gcagcagttg ttacagtttg ccaaggaaat cagtgtgagg  
8581 gcaaatacag aaaactggtc taaagaaacc ctgcaaagta cagttgacat ccttctccat  
8641 cacatattcc aagaacgaat ggatttggtt tctggaaatt ttctgaaaga actgtcctta  
8701 ataccattct tgtgtcctga acgggcccc gctgagtaca ttcggtttca ccctcagtac  
8761 caggaggtaa acggaacact tcctcttata aagttcaatg gagcacaagt gaatccaaag  
8821 ttcaagcaat gtgatgtact ccagctgctg tggacatctt gccctattct tccagagaaa  
8881 gccacaccgt tgagcattaa agaacaagaa ggcagtgacc tcgctccaca ggaacagctt  
8941 gaacaagttt taaatatgct taatgttaac ctggaccccc ctcttgataa ggtcattaat  
9001 aattgcagaa acatatgcaa cataacaact ttggatgagg aaatggtaaa aactagagca  
9061 aaggtcctaa ggagcatata tgaatttctg agtgcagaaa aacgagagtt ccgttttcag  
9121 cttcgggggtg tggcctttgt aatggtagaa gacggatgga aacttctgaa gcctgaggaa  
9181 gtagtgataa acctggagta tgaggctgat tttaaacctt atctgtacaa gctgccttta  
9241 gagcttggca cttttcatca gctgttcaaa catttaggta ctgaagatat catttccact  
9301 aagcaatatg ttgaagtgtt aagccgaata ttcaaaagct ctgaaggaaa gcagctagac  
9361 cctaatagaa tgcgtacagt taagagagtg gtttctggcc tattcaagag ttacaaaaat  
9421 gattcagtc aagtgaggag tgacctggag aatgccggg acctcgact ctacctcca  
9481 agccaggatg ggaagtgtgt gaagtcaagc atcttggtgt tcgatgatgc gccacattat  
9541 aaaagtagga tccaggggaa tattggcgtg cagatgctag ttgatcttag ccagtgtac  
9601 ttaggggaaag accatggatt tcacactaag ctgataatgc tctttctca aaagcttcga  
9661 cctcgtctgc tgagcagtat acttgaagag cagcttgatg aggagacccc taaagtgtgc  
9721 cagtttggcg cattgtgctc tcttcaggga agactgcagc ttctctgtc ttcagagcag  
9781 ttcattcacag gactcattcg aatcatgaag catgaaaatg ataatgctt cctggccaat  
9841 gaagaaaaag ccataagact ttgcaaagct ctaagagaag ggctgaaagt ttcctgtttt  
9901 gagaagcttc agacaacatt aagggttaaa ggttttaatc ctattcccca tagcaggagt  
9961 gaaactttcg cttttctaaa gcgatttggc aatgcagtca tcttgctcta catccaacat  
10021 tcagacagca aagacattaa ctttctgcta gccttagcga tgacacttaa atcagcaact  
10081 gacaatttga tttctgacac gtcatactta attgctatgc tgggatgcaa tgacatttac

Figure 8F

10141 aggatcagtg agaagcttga cagtttaggg gtgaaatagc actcctctga gccatcaaaa  
 10201 ctggaactcc ccatgcctgg cacaccaata cccgctgaga tccattacac actacttatg  
 10261 gatccaatga atgtttttta tcctggggaa tatgttggtt accttggtga tgctgaaggt  
 10321 ggtgatattc atgggtcata ccagccaaca tacacatacg caattattgt gcaagaagtt  
 10381 gaaagagaag atgctgacaa tactagtttc ttaggaaaga tctatcagat cgatattggc  
 10441 tacagtgaat ataagatagt cagctctctt gatctgtaca agttctcaag gcctgatgaa  
 10501 agctcccaaa acagagacag tgctcccacc acaccaacaa gccccaccga attcctgact  
 10561 cctggtctga gaagcatccc tcctcttttc tctggcaagg agagccacaa gtctccctcc  
 10621 accaaacacc attccccag aaagctcaag gtgaatgctt taccagaaat cttaaaagaa  
 10681 gtgacatcag tgggtggagca agcttggaag cttccagaat cagagcggaa aaagatcatt  
 10741 agacgcttgt atttgaagtg gcaccctgac aaaaatccag aaaatcatga tattgctaatt  
 10801 gaagtgttca agcacctgca gaatgaaatc aacagattag aaaaacaggc ttttctggat  
 10861 caaatgcag acagagcttc aagaagaaca ttttcaacct ctgcatctcg atttcagtca  
 10921 gacaagtact catttcaaag attttacact tcgtggaatc aagaagccac aagtcataaa  
 10981 tctgaaaggc aacagcaaag caaagagaaa tgccctcctt ctgctggaca gacatactct  
 11041 caaagggttct ttgttcctcc caccttcaag tcagtgggca atccagtggga agcccgagga  
 11101 tggttaagac aagccagagc aaacttctca gctgccagga atgaccttca caaaaatgcc  
 11161 aatgaatggg tgtgcttcaa gtgttacctt tccaccaagc tggctttgat tgcagccgac  
 11221 tatgctgtca gggggaaatc tgataaagat gtaaagccaa ctgcacttgc aaaaagata  
 11281 gaggagtaca gtcagcagct ggaaggactg acaaacgatg tgcacacatt ggaagcttat  
 11341 ggtgtagaca gcttgaaaac aaggtagcct gatttgcttc cttttccgca gattcccaat  
 11401 gacagggttca catctgaggt tgccatgagg gtgatggaat gcactgcctg tatcatcata  
 11461 aaacttgaat attttatata acagaaggtg tga

//

Figure 8G



LOCUS AF193556 12793 bp DNA PRI  
 DEFINITION Homo sapiens saccin (SACS) gene, complete cds.  
 ACCESSION AF193556 VERSION AF193556.1 GI:6907041KEYWORDS .  
 SOURCE human.  
 ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;  
 Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 12793) AUTHORS Engert,J.C., Berube,P., Mercier,J.,  
 Dore,C., Lepage,P., Ge,B., Bouchard,J.P., Mathieu,J., Melancon,S.B.,  
 Schalling,M., Lander,E.S., Morgan,K., Hudson,T.J. and Richter,A. TITLE ARSACS,  
 a spastic ataxia common in northeastern Quebec, is caused by mutations in a  
 new gene encoding an 11.5-kb ORF JOURNAL Nat. Genet. 24 (2), 120-125 (2000)  
 MEDLINE 20120709

REFERENCE 2 (bases 1 to 12793) AUTHORS Engert,J.C., Berube,P., Dore,C.,  
 Lepage,P., Ge,B., Hudson,T.J. and Richter,A. TITLE Direct Submission JOURNAL  
 Submitted (08-OCT-1999) Genome Centre, Montreal General Hospital, 1650 Cedar  
 Ave., Montreal, QC H3G 1A4, CanadaFEATURES Location/Qualifiers source 1..12793  
 /organism="Homo sapiens" /db\_xref="taxon:9606" /chromosome="13" /map="between  
 D13S232 and D13S292" mRNA 1..12793 /gene="SACS" /product="saccin" gene  
 1..12793 /gene="SACS" CDS 77..11566 /gene="SACS" /note="molecular chaperone"  
 /codon\_start=1 /product="saccin" /protein\_id="AAF31262.1"  
 /db\_xref="GI:6907042" /translation="

MNTFWPGRELIVQWYPFDENRNHPSVSWLKMVWKNLYIHFSEDL  
 TLFDEMLIPRTILEEGQTCVELIRLRIPSLVILDDSEEAQLPEFLADIVQKLGGFVL  
 KKLDASIQHPLIKKYIHSPLPSAVLQIMEKMPLQKLCNQITSLLPTHKDARLKFLASL  
 TDSSEKEKRRIQELAIKFRINHSSDQGISSYTKLKGCKVLHHTAKLPADLRLSISVID  
 SSDEATIRLANMLKIEQLKTTSCCLKVLKDIENAFYSHEEVTQLMLWVLENLSSLKNE  
 NPNVLEWLTPCLKFIQISQEQMVSAGELFDPDIEVLKDLFCNEEGTYFPSPSVFTSPDIL  
 HSLRQIGLKNEASLKEKDVVQVAKKIEALQVGACPDQDVLLKAKTLLLVLNKNHTLL  
 QSSEGKMTLKKIKWVPACKERPPNYPGSLVWKGDLNLCAPPDMCDVGHAILIGSSLP  
 LVESIHVNLEKALGIFTKPSLSAVLKHFKIVVDWYSSKTFSDDEYYQFHILLEIYGF  
 MHDHLNEGKDSFRALKFPWWWTGKKFCPLAQAVIKPIHDLDLQPYLHNPVKTMKFHQ  
 LFKVCGSIEELTSDHISMVQKIYKSDQDLSEQESKQNLHMLNIRWLYSNQIPAS  
 PNTVPVPIHHSKNPSKLIMKPIHECCYCDIKVDDLNDLLEDSEPIILVHEDIPMKTAE  
 WLKVPCLSTRLINPENMGFEQSGQREPLTVRIKNILEEYPSVSDIFKELLQONADDANA  
 TECSFLIDMRRNMDIRENLLDPGMAACHGPALWSFNNSQFSDSDFVNITRLGESLKR  
 EVDKVGKFGGLGFNSVYHITDIPIMSREFMIMFDPNINHISKHIKDKSNPGIKINWSK  
 QQKRLRKFPNQFKPFIDVFGCQLPLTVEAPYSYNGTLFRLSFRTQOEAKVSEVSSTCY  
 NTADIYSLVDEFSLCGHRLLIFTQSVKSMYKLYKIEETNPSLAQDTVIIKKKSCSSK  
 ALNTPVLSVLKEAAKLMKTSSSNKKLPSDEPKSSCILQITVEEFHHVFRRDIADLQSP  
 LFRGPDDPAALFEMAKSGQSKKPSDELSQKTVECTTWLLCTCMDTGEALKFSLSESG  
 RRLGLVPCGAVGVQLSEIQDQKWTVKPHIGEVFCYLPRIKTGLPVHINGCFAVTSNR  
 KEIWKTDTKGRWNTTFMRHVIVKAYLQVLSVLRDLATSGELMDYTYAVWPDPLVHD  
 DFSVICQGFYEDIAHGKGKELTKVFSKGSTWVSMKNVRFLDDSILKRRDVGSAAFKIF  
 LKYLKKTGSKNLCAVELPSSVKLGFEAGCKQILLENTFSEKQFFSEVFFPNIQEIEA  
 ELRDPLMIFVLNEKVDEFSGVLRVTPCIPCSLEGHPLVLPRLIHPEGRVAKLFDIKD  
 GRFPYGSTQDYLNPILIKLVQLGMAKDDILWDDMLERAVSVAEINKSDHVAACLRSS  
 ILLSLIDEKLRKIRDPRAKDFAAKYQTIKFLPFLTKPAGFSLDWKGNFSPKPMFAATD  
 LYTAHQDIVCLLQPIILNENSHSFRGCGSVSLAVKEFLGLLKKPTVDLVINQLKEVAK  
 SVDDGITLYQENITNACYKYLHEALMQNEITKMSIIDKLKPFSLVENAYVDSEKVS  
 FHLNFEAAPYLYQLPNKYKNNFRELFTVGVRSCTVEDFALVLESIDQERGKQITE  
 ENFQLCRRIISEGIWLSLIREKKQEFCEKNYKILLPDTNLMLLPAKSLCYNDPCWIKV

Figure 9A

KDTTVKYCHADIPREVAVKLGAVPKRHKALERYASNVCFTTLGTEFGQKEKLTSRIS  
 ILNAYPSEKEMLKELLQNADDAKATEICFVFDPRQHPVDRI FDDKWAPLQGPALCVYN  
 NQPFTEDDVRGIQNLGKGTKEGNPYKTGQYGIGFNSVYHITDCPSFISGNDILCIFDP  
 HARYAPGATSISPGRMFRDLADFRTOFSDVLDLYLGTHFKLDNCTMFRFPLRNAEMA  
 KVSEISSVPASDRMVQNLLDKLRSDGAELLMFLNHMEKISICEIDKSTGALNVLYSVK  
 GKITDGDRLKRQFHASVIDSVTKKRQLKDIPVQQITYTMDTEDSEGNLTWILICNRS  
 GFSSMEKVSXSIVSAHKNQDITLFPRGVAAACITHNYKKPHRAFCFLPLSLETGLPFH  
 VNGHFALDSARRNLWRDDNGVGRSDWNNSLMTALIPAYVELLIQLKKRYFPGSDPT  
 LSVLQNTPIHVVKDTLKKFLSFFPVNRLDLQPDLYCLVKALYNCIHEDMKRLLPVVRA  
 PNIDGSDLHSAVITWINMSTSNKTRPFFDNLLQDELQHLKNADYNITTRKTVAENVY  
 RLKHLLEIGFNLVYNCDATANLYHCLIDADIPVSYPVPADIRSFMTFSSPDTNCHI  
 GKLPCLRLQQTNLKLFHSLKLLVDYCFKDAEENEIEVEGLPLLITLDSVLQTFDAKRPK  
 FLTTYHELIPSRKDLFMENTLYLKYSNILLNCKVAKVFDISSFADLLSSVLPREYKTKS  
 CTKWKDNFASESWLKNAWHFISESVSVKEDQEETKPTFDIVVDTLKD WALLPGTKFTV  
 SANQLVVPEDGVLLPLSLMHIAVFPNAQSDKVFHALMKAGCIQLALNKICSKDSAFVP  
 LLSCHTANIESPTSILKALHYMVQTSTFRAEKLVENDEALLMYFNCNLNHLMSQDDI  
 KILKSLPCYKSISGRYVSIGKFGTCYVLTKSIPSAEVEKWTQSSSAFLEEKIHLKEL  
 YEYIGCVPVDDLEVLYLHLLPKIENLSYDAKLEHLIYLKNRLSSAEELSEIKEQLFEK  
 LESLLIHDANSRLKQAKHFYDRTVRVFEVMLPEKLFIPNDFFKLEQLIKPNHVT  
 MTSWVEFLRNIGLYILSQQLLQFAKEISVRANTENWSKETLQNTVDILLHHIFQER  
 MDLLSGNFLKELSLIPFLCPEPAPAEFIRFHPQYQEVNGTLPLIKFNGAQVNP KFKQC  
 DVLQLLWTSCPIPEKATPLSIKEQEGSDLGPQEQLQVLNMLNVNLDPPLDKVINNC  
 RNICNITTLDEEMVKTRAKVLRSIYEFLSAEKREFRQLRGVAFVMVEDGWKLLKPEE  
 VVINLEYESDFKPYLYKLPLELGTFFHQLFKHLGTEDIISTKQYVEVLSRIFKNSEKQ  
 LDPNEMRTVKRVVSGLFRSLQND SVKVRSDLENVRDLALYLP SQDGRLVKSSILVFDD  
 APHYKSRIQGNIGVQMLVDLSQCYLGKDHGFHTKLIMLPQKLRPRLSSILEEQ LDE  
 ETPKVCQFGALCSLQGRQLLLSSEQFITGLIRIMKHENDNAFLANEKAI RLCALR  
 EGLKVSCFEKLQTTLRVKGFNP IPHSRSETFAFLKRFGNVILLYIQHSDSKDINFL  
 ALAMTLKSATDNLISDTSYLIAMLGNDIYRIGEKLD SLGVKYSSEPSKLELPMPGT  
 PIPAEIHYTLLMDPMNVFPYGEYVGYLVDAEGGDIYGSYQPTYTYAIIVQEVEREDAD  
 NSSFLGKIYQIDIGYSEYKIVSSLDLYKFSRPEESSQSRDSAPSTPTSPTEFLTPGLR  
 SIPPLFSGRESHKTSSKHQSPKKLVNSLPEILKEVTSVVEQAWKL PESERKKIIRRL  
 YLKWHPDKNPENHDIANEVFKHLQNEINRLEKQAFLDQ NADRASRRTFSTSASRFQSD  
 KYSFORFYTSWNQEATSHKSERQQQNKCEKPPSAGQTY SQRFFVPPTFKSVGNPVEAR  
 RWLRQARANFSAARNDLHKANNEWVCFKCYLSTKLALIAADYAVRGKSDKDVKPTALA  
 QKIEEYSQLEGLTNDVHTLEAYGVDSLKTRYPDLLPFPQIPNDRFTSEVAMRVMECT  
 ACIIIKLENFMQQKV"

BASE COUNT 4163 a 2256 c 2487 g 3887 tORIGIN

atgatttaca ggaagaccat gtactcagct gcagcttcta aatccagaac gatttgcacg  
 tcttatcaag gaagtaatga atacattctg gcctggcaga gaattgattg ttcaatggta  
 tccatttgat gaaaacagaa atcacccatc tgtttcatgg cttaagatgg tttggaaaaa  
 tctttatata catttttcag aggatttgac tttatttgat gagatgccac ttatccccag  
 aactatacta gaggaaggtc agacatgtgt ggaactcatt agactcagga ttccatcggt  
 agtcatttta gacgatgaat ctgaagcaca gcttccagaa ttttttagcag acattgtaca  
 aaaacttgga ggggttggtc ttaaaaaatt agatgcatct atacaacatc cgcttattaa  
 aaaatatatt cattcaccat taccaagtgc tgttttgagc ataattggaga agatgccatt  
 gcagaaattg tgtaatcaaa taacttcgct acttccaaca cacaaagatg ccctgaggaa  
 gttcttggtc agtttaaccg atagcagtga gaaagagaaa agaattattc aagaattggc  
 aatattcaag cgcattaacc attcttctga tcagggaatt tcctcttata caaaattgaa  
 aggttgtaaa gtcttacacc atactgccaa actcccagca gatctgagac tttctatttc  
 agtaatagac agtagtgatg aagctactat tcgtctggca aacatgttga aaatagaaca  
 gttaaagacc actagctgct taaagcttgt tttaaaagat attgaaaatg cattttattc  
 acatgaagag gtaacacagc ttatgttatg ggtccttgag aatctatctt ctcttaaaaa

Figure 9B

tgagaatcca aatgtgcttg agtgggttaac accattaaaa ttcattccaga tatcacagga  
 acagatggta tcagctgggtg aactcttttga ccttgatata gaagtactaa aggatctctt  
 ttgtaatgaa gaaggaacct atttcccacc ctcagttttt acctaccag atattctttca  
 ctctttaaga cagattgggtt taaaaaacga agccagtctc aaagaaaagg atgttggtgca  
 agtggcaaaa aaaattgaag ccttacagggt cgggtgcttg cctgatcaag atgttcttct  
 gaagaaagcc aaaacctctt tactggtttt aaataagaat cacacactgt tgcaatcatc  
 tgaaggaaag atgacattga agaaaataaa atgggttcca gctgcaagg aaaggcctcc  
 aaattatcca ggctcttttg tctggaaagg agatctctgt aatctctgtg caccaccaga  
 tatgtgtgat gtaggccaatg caattctcat tggctcctca cttcctcttg ttgaaagtat  
 ccatgtaaac ctggaaaaag cattagggat cttcacaaaa cctagcctta gtgctgtctt  
 aaaacacttt aaaattgttg ttgattggta ttcttcaaaa accttagtg atgaagacta  
 ctatcaattc cagcatattt tgcttgagat ttacggattc atgcatgac atctaaatga  
 agggaaagat tcttttagag ccttaaaatt tccatgggtt tggactggca aaaagttttg  
 tccacttgcc caggctgtga ttaaaccaat ccatgatctt gaccttcagc cttatttgca  
 taatgtacct aaaacctggt caaaattcca ccaactattt aaggctctgtg gttcaataga  
 ggagttgaca tcagatcata ttcccatggt tattcagaag atatatctca aaagtgacca  
 agatctcagt gaacaagaaa gcaaacaaaa tcttcatctt atgttgaata ttatcagatg  
 gctgtatagc aatcagattc cagcaagccc caacacacca gttcctatac atcatagcaa  
 aaatccttct aaacttatca tgaagccaat tcacgaatgc tgttattgtg acattaaagt  
 tgatgacctt aatgacttac ttgaagattc tgtggaacca atcatttttg tgcattgagga  
 catacccatg aaaactgcag aatggctaaa agttccatgc cttagtacaa gactgataaa  
 tctgaaaac atgggatttg agcagtcagg acaaagagag ccacttactg taagaattaa  
 aaatattctg gaagaatacc cttcagtgct agatattttt aaagaactac ttcaaaacgc  
 tgatgatgca aatgcaacag aatgcagttt cttgattgat atgagaagaa atatggacat  
 aagagagaat ctcttagacc cagggatggc agcttgatc atggaacata actaggttag gagaatctt  
 caacaattct caattctcag attcagattt tgtgaacata actaggttag gagaatctt  
 aaaaagggga gaagttgaca aagttggaaa atttggtctt ggatttaatt ctgtgtacca  
 tatcactgac attcccatca ttatgagtcg ggaattcatg ataatgttcg atccaaacat  
 aaatcatatc agtaaacaca ttaaagacaa atccaatcct gggatcaaaa ttaattggag  
 taaacaacag aaaagactta gaaaatttcc taatcagttc aaaccattta tagatgtatt  
 tggctgtcag ttacctttga ctgtagaagc accttacagc tataatggaa cctttttccg  
 actgtccttt agaactcaac aggaagcaaa agtgagtga gttagtagta cgtgctacaa  
 tacagcagat atttattctc ttgtggatga atttagtctc tgtggacaca ggcttatcat  
 tttcactcag agtgtaaagt caatgtattt gaagtacttg aaaattgagg aaaccaaccc  
 cagtttagca caagatacag taataattaa aaaaaaatcc tgccttcca aagcattgaa  
 cacacctgtc ttaagtgttt taaaagaggc tgctaagctc atgaagactt gcagcagcag  
 taataaaaaag cttcccagtg atgaacaaa gtcattctgc attcttcaga tcacagtgga  
 agaatttcac catgtgttca gaaggattgc tgatttacag tcgccacttt ttagagggtc  
 agatgatgac ccagctgtc tctttgaaat ggctaagttc ggccaatcaa aaaagccatc  
 agatgagttg tcacagaaaa cagtagagtg taccacgtgg cttctgtgta cttgcatgga  
 cacaggagag gctctgaagt tttccctgag tgagagtgga agaagactag gactgggtcc  
 atgtggggca gtaggagttc agctgtcaga aatccaggac cagaagtgga cagtgaacc  
 acacattgga gaggtgtttt gctattttacc tttacgaata aaaacaggct tgccagttca  
 tatcaatggg tgctttgctg ttacatcaaa taggaaagaa atctggaaaa cagatacaaa  
 aggacgatgg aataccacgt tcatgagaca tgttattgtg aaagcttact tacagggtact  
 gagtgtctta cgggacctgg ccactagtgg ggagctaag gattatactt actatgcagt  
 atggcccgat cctgatttag ttcatgatga tttttctgta atttgccaa gattttatga  
 agatatagct catggaaaag ggaaagaact gaccaaagtc ttctctgatg gatctacttg  
 ggtttccatg aagaacgtaa gatttctaga tgactctata cttaaaagaa gagatgttgg  
 ttcagcagcc ttcaagatat ttttgaaata cctcaagaag actgggtcca aaaacctttg  
 tgctgttgaa cttccttctt cggtaaaaatt aggatgtgaa gaagctgggt gcaaacagat  
 actacttgaa aacacatttt cagagaaaca gttttttct gaagtgtttt ttccaaatat  
 tcaagaaatt gaagcagaac ttagagatcc tttaatgac tttgttctaa atgaaaaagt  
 tgatgagttc tcgggagttc ttcgtgttac tccatgtatt ccttgttcct tggaggggca

Figure 9C

tcctttgggtt	ttgccatcaa	gattgatcca	ccccgaagga	cgagttgcaa	agttatttga
tattaaagat	gggagattcc	cttatgggtc	tactcaggat	tatctcaatc	ctattatttt
gattaaacta	gttcagttag	gtatggcaaa	agatgatatt	ttatgggatg	atatgctaga
acgtgcagtg	tcagtagctg	aaattaataa	aagtgatcat	gttgctgcat	gcctaagaag
tagtatctta	ttgagtctta	tcgatgagaa	actaaaaata	agggatccta	gagcaaagga
ttttgctgca	aaatatcaaa	caatccgctt	ccttccattt	ctgacaaaac	cagcagggtt
ttctttggac	tggaaaaggca	acagttttta	gcctgaaacc	atgtttgcag	caactgacct
ttatacagct	gaacatcaag	atatagtttg	tcttttgcaa	ccaattctaa	atgaaaattc
ccattctttt	agaggtttgt	gttcagtgtc	attggctgtt	aaagagtttt	tgggattact
caagaagcca	acagttgatc	tggttataaa	ccaattgaaa	gaagtagcaa	aatcagttga
tgatgggaatt	acactgtacc	aggagaatat	caccaatgct	tgctacaaat	accttcacga
agccttgatg	caaaatgaaa	tcactaagat	gtcaattatt	gataagttaa	aaccctttag
cttcattcta	gttgagaatg	catatgttga	ctcagaaaag	gtttcttttc	atttaaattt
tgaggcggca	ccataccttt	atcagttgcc	taataagtat	aaaaataatt	tccgcgaact
ttttgaaacc	gtgggtgtga	ggcagtcagt	cactgttgaa	gattttgctc	ttgttttggg
atctattgat	caagaaaagag	gaacaaaagca	aataacagaa	gagaattttc	agctttgccg
acgaataatc	agtgaaggaa	tatggagtct	cattagagaa	aagaaaacaag	aattttgtga
gaaaaattat	ggcaagatat	tattgccaga	tactaatctt	atgcttctcc	ctgctaaatc
gttatgctac	aatgattgcc	cttggataaa	agtaaaggat	accactgtaa	aatattgtca
tgctgacata	cccagggaag	tagcagtaaa	actaggagca	gtcccaaagc	gacacaaaagc
cttagaaaaga	tatgcatcca	atgtctgttt	tacaacactt	ggcacagaat	ttgggcagaa
agaaaaattg	accagcagaa	ttaagagcat	ccttaatgca	tatccttctg	aaaaggaaat
gttgaaaagag	cttcttcaaa	atgctgatga	tgcaaaggcg	acagaaatct	gttttgtgtt
tgatcctaga	cagcatccag	ttgatagaat	atttgatgat	aagtgggccc	cattgcaagg
gccagcactt	tgtgtgtaca	acaaccagcc	atttacagaa	gatgatgtta	gaggaattca
gaatcttgga	aaaggcacga	aagagggaaa	tccttataaa	actggacagt	atggaatagg
attcaattct	gtgtatcata	tcacagactg	cccatctttt	atttctggca	atgacatcct
gtgtattttt	gatcctcatg	ccagatatgc	accagggggc	acatccatta	gtcccggagc
catgtttaga	gattttggatg	cagatttttag	gacacagttc	tcagatgttc	tggatcttta
tctgggaacc	catttttaaac	tggataattg	cacaatgttc	agatttcctc	ttcgtaatgc
agaaatggca	aaagtttcgg	aaatttcgtc	tgttccagca	tcagacagaa	tgggtccagaa
tcttttggac	aaactgcgct	cagatggggc	agaacttcta	atgtttctta	atcacatgga
aaaaatttct	atttgtgaaa	tagataagag	tactggagct	ctaaatgtgc	tgtattcagt
aaagggcaaa	atcacagatg	gagacagatt	gaaaaggaaa	caatttcacg	catctgtaat
tgatagtgtt	actaaaaaga	ggcagctcaa	agacatacca	gttcaacaaa	taacctatac
tatggatact	gaggactctg	aaggaaatct	tactacgtgg	ctaatttgta	atagatcagg
cttttcaagt	atggagaaaag	tatctaaaag	tgtcatatca	gctcacaaga	accaagatat
tactcttttc	ccacgtgggtg	gagtagctgc	ctgcattact	cacaactata	aaaaacccca
tagggccttc	tgttttttgc	ctctttcttt	ggagactggg	ctgccatttc	atgtgaatgg
ccactttgca	ctggattcag	ccagaaggaa	cctgtggcgt	gatgataatg	gagttgggtg
tcgaagtgac	tgggaataaca	gtttaatgac	agcattaata	gctcctgcat	atgttgaatt
gctaatacag	ttaaaaaaac	ggtatttccc	tggttctgat	ccaacattat	cagtgttaca
gaacacccct	attcatgttg	taaaggacac	tttaaagaag	tttttatcgt	ttttcccagt
taaccgtctt	gatctacagc	cagattttata	ttgtctagtg	aaagcacttt	acaattgcat
tcacgaagac	atgaaacgtc	ttttacctgt	tgtgcgggct	ccaaatattg	atggctctga
cttgactctt	gcagttataa	ttacttggat	caatatgtct	acttctaata	aaactagacc
attttttgac	aatttactac	aggatgaatt	acaacacctt	aaaaatgcag	attataatat
caccacacgc	aaaacagtag	cagagaatgt	ctataggctg	aaacatctcc	ttttagaaat
tggtttcaac	ttggtttata	actgtgatga	aactgcta	ctttaccact	gtcttataga
tgcatatatt	cctgttagtt	atgtgacccc	tgctgatata	agatcttttt	taatgacatt
ttcctctcct	gacactaatt	gccatattgg	gaagctgcct	tgtcgtctgc	agcagactaa
tctaaaactt	tttcatagtt	taaaactttt	agttgattat	tgttttaaag	atgcagaaga
aaatgagatt	gaagttgagg	gattgcccct	tctcatcaca	ctggacagtg	ttttgcaaac
ttttgatgca	aaacgaccca	agtttcta	aacatatcat	gaattgattc	catcccgc

Figure 9D

agacttggtt	atgaatacat	tatatattgaa	atatagtaat	atatttattga	actgttaaagt
tgcaaaaagt	tttgacattt	ccagctttgc	tgattttgtta	tcctctgtgt	tgccctcgaga
atataagacc	aaaagttgca	caaagtggaa	agacaatttt	gcaagtgagt	cttggccttaa
gaatgcatgg	cattttatta	gtgaatctgt	aagtgtgaaa	gaagatcagg	aagaaacaaa
accaacattt	gacattgttg	ttgatactct	aaaagactgg	gcattgcttc	caggaacaaa
gtttactggt	tcagccaacc	agcttgtggg	tcctgaagga	gatgttctgc	ttcctctcag
ccttatgcac	attgcagttt	ttccaaatgc	ccagagtgat	aaagtttttc	atgctctaat
gaaagccggc	tgtattcagc	ttgctttgaa	caaaatctgt	tccaaagaca	gtgcatttgt
tcctttgttg	tcatgtcaca	cagcaaatat	agagagcccc	acaagcatct	tgaaggctct
acattatatg	gtccaaactt	caacatttag	agcagaaaaa	ttagtagaaa	atgattttga
ggcacttttg	atgtatttca	actgcaattt	gaatcatttg	atgtcccaag	atgatataaa
aattctaaag	tcacttccgt	gctataaatc	catcagtggc	cgctatgtaa	gcattggaaa
atttggaaca	tgctacgtac	ttacaaaaag	tatcccttca	gctgaagtgg	agaaatggac
acaatcatca	tcactctgat	ttcttgaaga	aaaaatacac	ttaaaagaac	tatatgaggt
gattggttgt	gtacctgtag	atgatcttga	ggatatattg	aaacacctct	taccaaaaaat
tgaaaatctc	tcttatgatg	caaaattaga	gcacttgatc	taccttaaga	atagattatc
aagtgtctgag	gaattatcag	agattaagga	acaacttttt	gaaaaactgg	aaagtttatt
gataatccat	gatgctaaca	gtagactaaa	gcaagcaaag	catttctatg	atagaactgt
gagagttttt	gaagttatgc	ttcctgaaaa	attgtttatt	cctaatagatt	tctttaagaa
attggaacaa	cttataaaaac	ccaaaaatca	tgttacattt	atgacatcct	gggtggaatt
cttaagaaat	attggactaa	aatacatact	ttctcagcag	cagttgttac	agtttgctaa
ggaaatcagt	gtgagggtca	atacagaaaa	ctggtccaaa	gaaacattgc	aaaaatacagt
tgatatcctt	ctgcatcata	tattccaaga	acgaatggat	ttgttatctg	gaaattttct
gaaagaacta	tctttaatac	cattcttatg	tcctgagcgg	gcccccgagg	aattcatttag
atttcatcct	caatatcaag	aggtaaattg	aacacttcct	cttataaagt	tcaatggagc
acaggtaaat	ccaaaattca	agcaatgtga	tgtactccag	ctgttatgga	catcctgccc
tattcttcca	gagaaagcta	cacctttaag	cattaaagaa	caagaaggta	gtgaccttgg
tcacacaagaa	cagcttgaac	aagttttaaa	tatgcttaat	gttaacctgg	atcctcctct
tgataaggta	atcaataaact	gcagaaacat	atgcaacata	acgacgttgg	atgaagaaat
ggtaaaaaact	agagcaaaaag	tottaaggag	catatatgaa	ttcctcagtg	cagaaaaaaag
ggaatttcgt	tttcagttgc	gaggggttgc	ttttgtgatg	gtagaagatg	gttggaact
tctgaagcct	gaggaggtag	tcataaacct	agaatatgaa	tctgatttta	aaccttat
gtacaagcta	cctttagaac	ttggcacatt	tcaccagttg	ttcaaact	taggtactga
agatattatt	tcaactaagc	aatatgttga	agtgttgagc	cgcatattta	aaaattctga
gggcaaacaa	ttagatccta	atgaaatgcg	tacagttaag	agagtagttt	ctggtctggt
caggagtcta	cagaatgatt	cagtcaagggt	gaggagtgat	ctcgagaatg	tacgagacct
tgcgctttac	ctcccaagcc	aggatggtag	attggtaaag	tcaagcatct	tagtgtttga
cgatgcgcca	cattataaaa	gtagaatcca	ggggaatatt	ggtgtgcaaa	tgtagtttga
tctcagccag	tgctacttag	ggaaagacca	tggatttcac	actaagttga	taatgctctt
tcctcaaaaa	ccttagacctc	gattattgag	cagtatactt	gaagaacaat	tagatgaaga
gactcccaaa	gtttgtcagt	ttggagcggt	gtgttctctt	caaggaagat	tgcagttact
cttgtcttct	gaacagttca	ttacaggact	gattagaatt	atgaagcatg	aaaatgataa
tgcttttctg	gccaatgaag	aaaaagccat	aagactttgc	aaagccctaa	gagaaggatt
gaaagtatcc	tgctttgaaa	agcttcaaac	aacattaaga	gttaaagggt	ttaatcctat
tcctcacagc	agaagtgaag	cttttgcttt	tttgaagcga	tttggtaatg	cagtcactct
gctctacatt	caacattcag	acagtaaaga	cattaatttc	ctgttagcac	tggcaatgac
tcttaaatca	gcaactgaca	atgtgatttc	tgacacttca	tatttaattg	ctatgctagg
atgcaatgat	atttacagga	ttgggtgaga	acttgacagt	ttaggagtga	aatatgactc
ttcggagcca	tcaaaactgg	aacttccaat	gcctggcaca	ccaattcctg	ctgaaattca
ttacactctg	cttatggacc	caatgaatgt	tttttaccg	ggagaatatg	ttgggtacct
tgttgatgct	gaaggtggtg	atatctatgg	atcataccag	ccaacataca	catatgcaat
tattgtacaa	gaagttgaaa	gagaagatgc	tgacaattct	agttttctag	gaaagatata
tcagatagat	attggttata	gtgaatataa	aatagttagc	tctcttgatc	tgtataagtt
ttcaagacct	gaggaaagct	ctcaaagcag	ggacagtgtc	ccttctacac	caaccagccc

Figure 9E

cactgagttc	ctcacccctg	gcttgagaag	cattcctcct	cttttctctg	gtagagagag
ccacaagact	tcttccaaac	atcagtcccc	caaaaagctt	aagggttaatt	ctttaccaga
aatcttaaaa	gaagtgacat	ctgtggtgga	gcaagcatgg	aagcttccag	aatcggaacg
aaaaaagatt	attaggcggt	tgtatttgaa	atggcatcct	gacaaaaatc	cagagaacca
tgacattgcc	aatgaagttt	ttaaacattt	gcagaatgaa	atcaacagat	tagaaaaaca
ggctttttcta	gatcaaaaatg	cagacagggc	ctccagacga	acattttcaa	cctcagcatc
ccgattttcag	tcagacaaaat	actcatttca	gagatttctat	acttcatgga	atcaagaagc
aacgagccat	aaatctgaaa	gacagcaaca	gaacaaagaa	aaatgcccc	cttcagccgg
acagacttac	tctcaaagggt	tctttgttcc	tcccactttc	aagtcgggtg	gcaatccagt
ggaagcacgc	agatggctaa	gacaagccag	agcaaacttc	tcagctgcca	ggaatgacct
tcataaaaaat	gccaatgagt	gggtgtgctt	taaatgttac	ctttctacca	agtttagcttt
gattgcagct	gactatgctg	tgaggggaaa	gtctgataaa	gatgtaaaac	caactgcact
tgctcagaaa	atagaggaat	atagtcagca	acttgaagga	ctgacaaatg	atgttcacac
attggaagct	tatggtgtag	acagttttaa	aacaagatac	cctgatttgc	ttccctttcc
tcagatccca	aatgacaggt	tcacttctga	ggttgctatg	aggggtgatgg	aatgtactgc
ctgtatcata	ataaaaacttg	aaaattttat	gcaacaaaaa	gtgtgaagat	atttaacgaa
aaaaaaggta	gatcttgaat	gtgttgtagc	acgaataaat	tgctgtactt	cattaagctt
cattgccaat	tagctaggaa	ttgttaagca	cattgcagat	tgttcttggg	gaattctgga
gttggttatga	acatgaatac	caacggaaaa	ccttaactga	atctaaaaga	aaactatttt
gaagatgggtg	gtgagctgca	aaatagctgg	atggatttga	atgattggga	tgatacatca
ttgaactgca	ctttatataa	ccaaagctta	gcagtttgtt	agataagagt	ctatgtatgt
ctctgggttag	gatgaagtta	attttatgtt	tttaacatgg	tatttttgaa	ggagctaattg
aaacactgga	catataattg	gtttaaacat	aaggggaatt	aagtccttgt	agtctgtcat
ttttttaagt	ggatcctctt	ggatgcgtta	ttttctcatc	agctggctct	gatcatgaat
ttgttgtaat	tttatgttgt	actcagtgca	tttaagaaat	ggtagagtat	tttaatccta
ttacttgact	aagagtgtga	aggtagtact	ttttagagtg	cactgagtgc	actttacatc
tttattttaa	ttttttttta	acatcttatg	tttacaggct	tcctgtttga	tgaagatagc
aacggaaaaac	tcaaaatgggt	ggcagttcct	attaccagtt	gttagtattg	tttctggaaa
ctgcttgcca	agacaacatt	tattaactgt	tagaacactt	gctttatgtt	tgtgtgtaca
tattttccac	aaatgttata	atttatatag	tgtggttgaa	caggatgcaa	tcttttgttg
tctaaagggtg	ctgcagttaa	aaaaaaaaaca	accttttctt	tcaatatggc	atgtagtgga
gttttttttaa	ctttaaaaaac	atcaaaaatt	gttaaaatca	ttgtgttatc	tagtagttta
taattatcgg	cttatatttc	cccatgaatg	atcagaactg	acatttaatt	catgtttgtc
tcgccatgct	tctttacttt	aacatatttc	ttttgcagaa	tgtaaaagggt	aatgataatt
agttttatata	agtgtactgg	ctgtaaaatga	tgctaaaatat	acttttatgca	attaagggct
tacagaacat	gttgaaaactt	tttttacttt	tattgggaat	aaggaatgtt	tgcacctcca
cattttattg	ctt				

Figure 9F